

Ad auctions and cascade model: GSP inefficiency and algorithms

Gabriele Farina and Nicola Gatti

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano
Piazza Leonardo da Vinci, 32
I-20133, Milan, Italy
gabriele2.farina@mail.polimi.it, nicola.gatti@polimi.it

Abstract

The design of the best *economic mechanism* for Sponsored Search Auctions (SSAs) is a central task in computational mechanism design/game theory. Two open questions concern the adoption of user models more accurate than that one currently used and the choice between Generalized Second Price auction (GSP) and Vickrey–Clark–Groves mechanism (VCG). In this paper, we provide some contributions to answer these questions. We study Price of Anarchy (PoA) and Price of Stability (PoS) over social welfare and auctioneer’s revenue of GSP w.r.t. the VCG when the users follow the famous *cascade model*. Furthermore, we provide exact, randomized, and approximate algorithms, showing that in real-world settings (Yahoo! Webscope A3 dataset, 10 available slots) optimal allocations can be found in less than 1s with up to 1,000 ads, and can be approximated in less than 20ms even with more than 1,000 ads with an average accuracy greater than 99%.

Introduction

SSAs, in which a number of *advertisers* bid to have their *ads* displayed in some slot alongside the search results of a keyword, constitute one of the most successful applications of *microeconomic mechanisms*, with a revenue of about \$50 billion dollars in the US alone in 2014 (IAB 2015). Similar models are used in many other advertisement applications, e.g. contextual advertising (Varian and Harris 2014). A number of questions remain currently open in the study of effective SSAs. Among these, two are crucial and concern the adoption of the best *user model* in real-world settings and the design of the best *economic mechanism*. In this paper, we provide some contributions to answer these questions.

The commonly adopted user model is the *position dependent cascade* (PDC) in which the user is assumed to observe the slots from the top to the bottom with probabilities to observe the next slot that depend only on the slots themselves (Narahari et al. 2009). The optimal allocation can be efficiently found in a greedy fashion. The natural extension is the *ad/position-dependent cascade* (APDC) model, in which the probability to observe the next slot depends also on the ads actually allocated. This model is

extensively experimentally validated (Craswell et al. 2008; Joachims et al. 2007). However, it is not known whether finding the best allocation of APDC is \mathcal{NP} -hard, but it is commonly conjectured to be. (Gatti and Rocco 2013) provides an exact algorithm that can be used in real time only with 40 ads or less, while in real-world applications the ads can be even more than 1,000. The problem admits a constant-ratio $(1 - \epsilon)/4$, with $\epsilon \in [0, 1]$, approximation algorithm (Gatti and Rocco 2013; Kempe and Mahdian 2008), but no known algorithm is proved to lead to a *truthful* economic mechanism. In addition, the known approximation algorithms can be applied in real time only with 100 ads or less (Gatti and Rocco 2013). More expressive models include a contextual graph, see (Gatti et al. 2015; Fotakis, Krysta, and Telelis 2011), but they are Poly- \mathcal{APX} -hard. Hence, the APDC is considered as the candidate providing the best trade-off between accuracy and tractability.

While the GSP is still popular in many SSAs, the increasing evidence of its limits is strongly pushing towards the adoption of the more appealing VCG mechanism, which is already successfully employed in the related scenario of contextual advertising, by Google (Varian and Harris 2014) and Facebook (Hegeman 2010). The main drawback of the GSP is the *inefficiency* of its equilibria (in terms of social welfare) w.r.t. the VCG outcome (the only known results concern PDC): considering the whole set of Nash equilibria in full information, the PoA of the GSP is upper bounded by about 1.6, while considering the set of Bayes–Nash equilibria the PoA is upper bounded by about 3.1 (Leme and E. Tardos 2010). Furthermore, automated bidding strategies, used in practice by the advertisers to find their best bids, may not even converge to any Nash equilibrium and, under mild assumptions, the states they converge to are shown to be arbitrarily inefficient (Evangelos and Telelis 2010). Some works study how inefficiency is affected by some form of externalities, showing that no Nash equilibrium of the GSP provides a larger revenue than the VCG outcome (Kuminov and Tennenholtz 2009; Gomes, Immorlica, and Markakis 2009).

Original contributions. Under the assumption that the users behave as prescribed by APDC, we study the PoA and PoS both over the social welfare and over the auctioneer’s revenue of the GSP w.r.t. the VCG based on APDC (we study also the restricted case in which the advertisers do

not overbid). We extend our analysis done for the GSP to the VCG based on PDC. Both analyses show a wide dispersion of the equilibria with PoAs in most cases unbounded, pointing out the limits of the traditional models. Furthermore, we provide a polynomial-time algorithm that dramatically shrinks auctions instances, safely discarding ads (even more than 90% in our experiments). We also provide a randomized algorithm (that can be derandomized with polynomial cost) finding optimal allocations with probability 0.5 in real-world instances (based on the *Yahoo!* Webscope A3 dataset) with up to 1,000 ads and 10 slots in less than 1s and a *maximal-in-its-range* (therefore leading to truthful mechanisms if paired with VCG-like payments) approximation algorithm returning allocations with an average approximation ratio larger than about 0.99 in less than 20ms even with 1,000 ads. This shows that the APDC can be effectively used in practice also in large instances. Finally, we prove our approximation algorithm to have a ratio of 1/2 in restricted settings. However, in general settings the proof keeps to be elusive (even if we did not find any instance in which the approximation ratio is below 1/2).

Problem formulation

We adopt the same SSA model used in (Gatti and Rocco 2013) with minor variations. The model without externalities is composed of the following elements:

- $\mathcal{A} = \{1, \dots, N\}$ is the set of ads. W.l.o.g. we assume each advertiser (also called agent) to have a single ad, so each agent $a \in \mathcal{A}$ can be identified with ad a ;
- $\mathcal{K} = \{1, \dots, K\}$ is the set of slots s ordered from the top to the bottom. We assume w.l.o.g. $K \leq N$;
- $q_a \in [0, 1]$ is the *quality* of ad a (i.e., the probability a user will click ad a once observed);
- $v_a \in V_a \subseteq \mathbb{R}^+$ is the value for agent a when ad a is clicked by a user, while $\mathbf{v} = (v_1, \dots, v_N)$ is the value profile;
- \bar{v}_a is the product of q_a and v_a , and can be interpreted as the expected fraction of the ad value that is actually absorbed by the users observing ad a ;
- $\hat{v}_a \in \hat{V}_a \subseteq \mathbb{R}^+$ is the value reported by agent a , while $\hat{\mathbf{v}} = (\hat{v}_1, \dots, \hat{v}_N)$ is the reported value profile;
- Θ is the set of (ordered) allocations of ads to slots, where each ad cannot be allocated in more than one slot.

Given an allocation θ of ads to slots, we let:

- $\text{ads}(\theta) \subseteq \mathcal{A}$ to denote the subset of ads allocated in θ ;
- $\text{slot}_\theta(a)$ to denote the slot in which ad a is allocated, if any;
- $\text{ad}_\theta(s)$ to denote the ad allocated in slot s .

The externalities introduced by the cascade model assume the user to have a Markovian behavior, starting to observe the slots from the first (i.e., slot 1) to the last (i.e., slot K) where the transition probability from slot s to slot $s+1$ is given by the product of two parameters:

- (*ad-dependent externalities*) $c_a \in [0, 1]$ is the *continuation probability* of ad a ;
- (*position-dependent externalities*) $\lambda_s \in [0, 1]$ is the *factorized prominence* of slot s (it is assumed $\lambda_K = 0$).

The click through rate $\text{CTR}_\theta(a) \in [0, 1]$ of ad a in allocation θ is the probability a user will click ad a and it is formally defined as

$$\text{CTR}_\theta(a) = q_a \cdot \Lambda_{\text{slot}_\theta(a)} \cdot C_\theta(a),$$

where

$$C_\theta(a) = \prod_{s < \text{slot}_\theta(a)} c_{\text{ad}_\theta(s)},$$

$$\Lambda_s = \prod_{k < s} \lambda_k.$$

Parameter Λ_s is commonly called *prominence* (Kempe and Mahdian 2008).

Given an allocation θ , its social welfare is defined as:

$$\text{SW}(\theta) = \sum_{\text{ads}(\theta)} v_a \cdot \text{CTR}_\theta(a) = \sum_{\text{ads}(\theta)} \bar{v}_a \cdot \Lambda_{\text{slot}_\theta(a)} \cdot C_\theta(a).$$

The problem we study is the design of an economic mechanism \mathcal{M} , composed of

- an *allocation function* $f : \times_{a \in \mathcal{A}} V_a \rightarrow \Theta$ and
- a *payment function* for each agent a , $p_a : \times_{a \in \mathcal{A}} \hat{V}_a \rightarrow \mathbb{R}$.

Each agent a has a linear utility $u_a(v_a, \hat{\mathbf{v}}) = v_a \cdot \text{CTR}_{f(\mathbf{v})}(a) - p_a(\hat{\mathbf{v}})$, in expectation over the clicks and pays $p_a(\hat{\mathbf{v}})/\text{CTR}_{f(\hat{\mathbf{v}})}(a)$ only once its ad is clicked. We are interested in mechanisms satisfying the following properties:

Definition 1. Mechanism \mathcal{M} is *dominant strategy incentive compatible (DSIC)* if reporting true values is a dominant strategy for every agent (i.e., $\hat{v}_a = v_a$).

Definition 2. Mechanism \mathcal{M} is *individually rational (IR)* if no agent acting truthfully prefers to abstain from participating to the mechanism rather than participating.

Definition 3. Mechanism \mathcal{M} is *weakly budget balance (WBB)* if the mechanism is never in deficit.

Definition 4. Mechanism \mathcal{M} is *computationally tractable* when both f and p_a are computable in polynomial time.

Definition 5. Allocation function f is *maximal in its range* if f returns an allocation maximizing the social welfare among a given subset of allocations that is independent of the agents' reports.

When f is maximal in its range, VCG-like payments can be used, obtaining a DSIC, IR and WBB mechanism (Nisan and Ronen 2000).

Inefficiency of GSP and VCG with PDC

GSP analysis

In the GSP, ads are allocated according to decreasing reported values: supposing $\hat{v}_1 \geq \dots \geq \hat{v}_N$, the allocation function maps ad a to slot $s = a$, for each $a = 1, \dots, K$. Every agent $a = 1, \dots, K-1$ is charged a price $p_a = q_{a+1}\hat{v}_{a+1}$. Agent K is charged a price $p_K = 0$ if $K = N$, or $p_K = q_{K+1}\hat{v}_{K+1}$ otherwise. The following lemma holds:

Lemma 1. *The GSP is not IR, when users follow the APDC.*

The GSP is well known not to be DSIC. For this reason, we study the inefficiency of its Nash equilibria.

Lemma 2. *The PoA of the social welfare in the GSP when users follow APDC is unbounded.*

Lemma 3. *The PoA of the social welfare in the GSP when users follow APDC is $\geq K$, in the restricted case the agents do not overbid.*

Lemma 4. *The PoA of the revenue in the GSP when users follow APDC is unbounded (even without overbidding).*

Lemma 5. *The PoS of the social welfare in the GSP when users follow APDC is 1 (even without overbidding).*

Lemma 6. *The PoS of the revenue in the GSP when users follow APDC is 0 (even without overbidding).*

Proofs of the previous lemmas can be found in the Appendix . Lemma 2 and 3 together show a huge dispersion of the social welfare of the allocations associated with GSP equilibria: it may be arbitrarily larger than the social welfare of the APDC. Lemma 4 and 6 show that this situation gets even worse when we turn our attention onto the auctioneer's revenue, with the revenue of GSP equilibria being arbitrarily smaller or larger than that of the APDC.

VCG with PDC analysis

The PDC model is very similar to the APDC: the difference lies in the fact that $c_a = 1$ for each ad $a \in \mathcal{A}$. Payments are calculated according to the rules attaining the VCG mechanism, thus providing a IR, WBB, DSIC mechanism under the assumption that users experience ad fatigue due only to the positions of the slots and not to the allocated ads. When the user's behavior is affected by ad continuation probabilities, the above mentioned properties do not necessarily apply anymore. Indeed, the following lemma holds:

Lemma 7. *The VCG with PDC is neither IR nor DSIC, when users follow APDC.*

We study the inefficiency of Nash equilibria, VCG with PDC not being truthful.

Lemma 8. *The PoA of the social welfare in the VCG with PDC when users follow APDC is unbounded.*

Lemma 9. *The PoA of the social welfare in the VCG with PDC when users follow APDC is $\geq K$, in the restricted case the agents do not overbid.*

Lemma 10. *The PoA of the revenue in the VCG with PDC when users follow APDC is unbounded (even without overbidding).*

Lemma 11. *The PoS of the social welfare in the VCG with PDC when users follow APDC is 1.*

Lemma 12. *The PoS of the revenue in the VCG with PDC when users follow APDC is 0.*

Lemma 13. *The PoS of the revenue in the VCG with PDC when users follow APDC is ≤ 1 , in the restricted case the agents do not overbid.*

Proofs of the previous lemmas can be found in the Appendix . Again, we see a huge dispersion, both in terms of social welfare and of revenue, among the different equilibria of VCG with PDC.

Algorithms

DOMINATED-ADS algorithm

We present an algorithm meant to reduce the size of the problem (i.e., the number of ads), without any loss in terms of social welfare or revenue of the optimal allocation. The central observation for our algorithm is that, under certain circumstances, given two ads a, b with parameters (\bar{v}_a, c_a) and (\bar{v}_b, c_b) respectively, it is possible to establish *a priori* that, if in an optimal allocation b is allocated to a slot, then ad a is allocated in a slot preceding that of b ; whenever this is the case we say that ad a “dominates” ad b . As an example, consider two ads a and b , satisfying the condition $(\bar{v}_a > \bar{v}_b) \wedge (c_a > c_b)$: in accordance with intuition, a simple exchange argument shows that a dominates b . A weaker sufficient condition for deciding whether a dominates b is given in the following lemma and proved in the Appendix .

Lemma 14. *Let $\lambda_{\max} = \max\{\lambda_s : s \in \mathcal{K}\}$, and B an upper bound¹ of the value of the optimal allocation; also, let $\mathcal{D} = [0, \lambda_{\max}] \times [0, B]$. Given two ads a, b with parameters (\bar{v}_a, c_a) and (\bar{v}_b, c_b) , consider the affine function $w_{a,b} : \mathcal{D} \rightarrow \mathbb{R}$, defined as*

$$w_{a,b}(x, y) = \det \begin{pmatrix} x & -y & 1 \\ 1 & \bar{v}_b & c_b \\ 1 & \bar{v}_a & c_a \end{pmatrix}$$

If the minimum of $w_{a,b}$ over \mathcal{D} is greater than 0, then a dominates b .

We will use the notation $a \prec b$ to denote that ad a dominates ad b , in the sense of Lemma 14. Note that \prec defines a partial order over the set of ads. Since $w_{a,b}$ is an affine function defined on a convex set, it must attain a minimum on one of the corner points of \mathcal{D} , hence the following holds:

Lemma 15. *If the four numbers $w_{a,b}(0, 0)$, $w_{a,b}(0, B)$, $w_{a,b}(\lambda_{\max}, 0)$, $w_{a,b}(\lambda_{\max}, B)$ are all positive, then $a \prec b$.*

We define the *dominators* of an ad a as the set $\text{dom}(a) = \{b \in \mathcal{A} : b \prec a\}$. The following lemma is central to our algorithm, as it expresses a sufficient condition for discarding an ad from the problem:

Lemma 16. *If $|\text{dom}(a)| \geq K$, then ad a can be discarded, as it will never be chosen for an allocation.*

This suggests a straightforward algorithm to determine the set of safely deletable ads, as streamlined in Algorithm 1. In the Appendix we show two methods, respectively named CONST- λ and DECOUPLE, for calculating a

¹ More precisely, it is enough that B is a (weak) upper bound of the quantity

$$\tilde{B} = \max_i \{\lambda_i \cdot \text{ALLOC}(i + 1, K)\},$$

where $\text{ALLOC}(i, K)$ is the value of the optimal allocation of the problem having as slots the set $\mathcal{K}_i = \{i, \dots, K\} \subseteq \mathcal{K}$.

Algorithm 1 DOMINATED-ADS

```
1: procedure DOMINATED-ADS(ads, slots)
2:   Determine  $\mathcal{D}$ , as defined in Lemma 14
3:   For each ad  $a$ , compute  $|\text{dom}(a)|$ 
4:   Discard all ads  $a$  having  $|\text{dom}(a)| \geq K$ 
```

“good” upper bound B that is needed in Line 2 in order to determine \mathcal{D} . The first method (CONST- λ) has a computational complexity of $O(NK)$ time, while DECOUPLE runs in $O(N + K \log K)$ time and makes use of the FFT algorithm (Cooley and Tukey 1965). Both methods require linear, i.e. $O(N + K) = O(N)$ memory. Of course, when K is small enough, both algorithms can be run and the smallest upper bound can be taken as the value of B .

A naïve implementation of Line 3 tests every ad a against all the other ads, keeping track of the number of dominators of a . This leads to an easily implementable $O(N^2)$ algorithm suitable for instances with $\approx 1,000$ ads, but is not applicable in real time when N grows. In the Appendix we show that it is possible to compute the number of dominators of an ad a in $O(\log N)$ amortized time using the dynamic fractional cascading technique (Mehlhorn and Näher 1990). Hence, the following lemma holds:

Lemma 17. *The DOMINATED-ADS algorithm runs in $O(N \log N)$ or $O(N(K + \log N))$ time, depending on the employed upper bounding strategy, and $O(N)$ memory.*

Finally notice that iterating the algorithm could lead to successive simplifications of the ads set, as the upper bound B provided by CONST- λ and DECOUPLE may decrease in successive runs of DOMINATED-ADS. However, the maximum number of iterations is N , since, if no ad has been discarded, then B does not change and no further ad can be discarded in future iterations.

COLORED-ADS algorithm

We present an algorithm that can determine an optimal allocation in time polynomial in N and exponential in K . As the number of slots is generally small (typically ≤ 10), this algorithm is suitable for real-world applications. The key insight for the algorithm is that the problem of finding the allocation maximizing the social welfare can be cast to an instance of the well-known problem of finding a maximal K -vertex weighted simple path in an un directed graph. Efficient algorithms for the latter problem can therefore be used to solve our problem.

We introduce the definition of a *right-aligned allocation*.

Definition 6. *Given an ordered sequence $S : a_1, \dots, a_n$ of ads, we say that the right-aligned allocation of S is the allocation mapping the ads in S to the last n slots, in order, leaving the first $K - n + 1$ slots vacant.*

Consider the complete undirected graph \mathcal{G} having the available ads for vertices; every simple path $\pi : a_1, \dots, a_n$

of length $n \leq K$ in \mathcal{G} can be uniquely mapped to the right-aligned allocation of π , and *vice versa*. Following this correspondence, we define the *value* of a simple path in \mathcal{G} as the value of the corresponding right-aligned allocation. In order to prove that the problem of finding a maximal K -vertex weighted simple path in \mathcal{G} is equivalent to that of finding an optimal allocation, it is sufficient to prove that there exists (at least) one optimal allocation leaving no slot vacant, i.e. using all of the K slots. To this end we introduce the following lemma, whose proof is in the Appendix .

Lemma 18. *In at least one optimal allocation all the available slots are allocated.*

The problem of finding a maximal K -vertex weighted simple path in \mathcal{G} can be solved by means of the *color coding* technique (Alon, Yuster, and Zwick 1994). The basic idea is as follows: in a single iteration, vertices are assigned one of K random colors; then, the best path visiting every color exactly once is found using a dynamic programming approach; finally, this loop is repeated a certain amount of times, depending on the type of algorithm (randomized or derandomized). In the randomized version, the probability of having missed the best K -path decreases as e^{-R/e^K} , where R is the number of iterations. Therefore, using randomization, we can implement a $O((2e)^K N)$ compute time and $O(2^K + N)$ memory space algorithm, as streamlined in Algorithm 2. In the derandomized version, the algorithm is exact and requires $O((2e)^K K^{O(\log K)} N \log N)$ time when paired with the derandomization technique presented in (Naor, Schulman, and Srinivasan 1995).

Algorithm 2 COLORED-ADS

```
1: procedure COLORED-ADS(ads, slots)
2:   repeat  $e^K \log 2$  times  $\triangleright$  50% success probability
3:     Assign random colors  $\in \{1, \dots, K\}$  to  $\mathcal{G}$ 's vertices
       note: each color must be used at least once

 $\triangleright$  We now construct a memoization table  $\text{MEMO}[\mathcal{C}]$ , reporting,
for each color subset  $\mathcal{C}$  of  $\{1, \dots, K\}$  the maximum value of
any simple path visiting colors in  $\mathcal{C}$  exactly once.

4:    $\text{MEMO}[\emptyset] \leftarrow 0$ 
5:    $\mathcal{P} \leftarrow$  powerset of  $\{1, \dots, K\}$ , sorted by set size
6:   for  $\mathcal{C} \in \mathcal{P} - \{\emptyset\}$ , in order do
7:      $\tilde{\lambda} \leftarrow \lambda_{K-|\mathcal{C}|+1}$ 
8:     for each color  $c \in \mathcal{C}$  do
9:       for each vertex (ad)  $a$  in  $\mathcal{G}$  of color  $c$  do
10:         $\text{VALUE} \leftarrow \bar{v}_a + c_a \tilde{\lambda} \cdot \text{MEMO}[\mathcal{C} - \{c\}]$ 
11:         $\text{MEMO}[\mathcal{C}] \leftarrow \max\{\text{MEMO}[\mathcal{C}], \text{VALUE}\}$ 
12:   return  $\text{MEMO}[\{1, \dots, K\}]$ 
```

Note that Algorithm 2 is just a simplified, randomized and non-parallelized sketch of the algorithm we test in the experimental section of this paper; also, for the sake of presentation, in Algorithm 2 we only return the value of the optimal allocation and not the allocation itself. We also considered the idea of using $1.3K$ different colors, as suggested by the work of (Hüffner, Wernicke, and Zichner 2008), but we found out that for this particular problem the improvement is negligible.

We conclude this subsection with some remarks about the above algorithm. First, we remark that, given the nature of the operations involved, the algorithm proves to be efficient in practice, with only a small constant hidden in the big-oh notation. Furthermore, it is worth to note that the iterations of the main loop (Lines 2 to 15) are independent; as such, the algorithm scales well horizontally in a parallel computing environment. Second, we point out an important economic property of the algorithm, that makes COLORED-ADS appealing: it allows the design of truthful mechanisms when paired with VCG-like payments. While this is obviously true when the algorithm is used in its derandomized exact form, it can be proven that the truthfulness property holds true even when only a partial number of iterations of the main loop is carried out. This easily follows from the fact that the algorithm is maximal-in-its-range, searching for the best allocation in a range that does not depend on the reports of the agent. This implies that it is possible to interrupt the algorithm after a time limit has been hit, without compromising the truthfulness of the mechanism. This leads to approximate algorithms that offer convenient time-approximation trade-offs, as long as K is small. We experimentally investigate the approximation ratio of these algorithms as a function of the number of performed iterations of the main loop in the Appendix .

SORTED-ADS approximate algorithm

While the general problem of finding an optimal allocation is difficult, polynomial time algorithms are easy to derive—as we show below—when we restrict the set of feasible allocations to those respecting a given total order \prec_{ads} defined on the ads set. This suggests this simple approximation algorithm: first, T total orders $\prec_{\text{ads},1}, \dots, \prec_{\text{ads},T}$ over the ads set are chosen; then, the optimal allocation satisfying $\prec_{\text{ads},i}$ is computed, for each i ; finally, the value of the optimal allocation for the original unrestricted problem is approximated with the best allocation found over all the T orders. The number of total orders T is arbitrary, with more orders obviously producing higher approximation ratios. An empirical study on the time-approximation tradeoff curve of SORTED-ADS is available in the Appendix .

In order to find the optimal allocation respecting the total order \prec_{ads} , we propose a simple $O(NK)$ time dynamic programming algorithm, which we name SORTED-ADS, described in Algorithm 3. The idea behind the algorithm is to find, for each $n = 1, \dots, N$ and $k = 1, \dots, K$, the value $\mathcal{T}[n, k]$ of the best allocation for the problem having $\mathcal{A}_n = \{1, \dots, n\} \subseteq \mathcal{A}$ as ads set and $\mathcal{K}_k = \{k, \dots, K\} \subseteq \mathcal{K}$ as slots set. The values of $\mathcal{T}[n, k]$ can be computed inductively, noticing that the associated subproblems share the same optimal substructure. As before, in Algorithm 3 we only show how to find the value of the optimal allocation, and not the allocation itself. Also, for the sake of simplicity, the algorithm we present uses $O(NK)$ memory, but we remark that it is possible to bring the required memory down to $O(N + K) = O(N)$ while letting fast (i.e. $O(NK)$ time) reconstruction of the optimal allocation, using a technique similar to that presented in (Hirschberg 1975). See the Appendix for further details.

Algorithm 3 SORTED-ADS

```

1: procedure SORTED-ADS(ads, slots,  $\prec_{\text{ads}}$ )
2:   Sort the ads according to  $\prec_{\text{ads}}$ , so that ad 1 is the minimum
   ad w.r.t. the given order
   ▷ We now construct a memoization table  $\mathcal{T}[n, k]$ , reporting,
   for each  $1 \leq n \leq N$  and  $1 \leq k \leq K$ , the value of the best
   allocation that uses only ads  $n, \dots, N$  and slots  $k, \dots, K$  (i.e.
   no ads get allocated to slots  $1, \dots, k-1$  and  $\lambda_i = 1, \forall i < k$ ).
3:    $\mathcal{T}[N, k] \leftarrow \bar{v}_N, \quad k = 1, \dots, K$  ▷ Base case
4:   for  $n = N - 1$  downto 1 do
5:     for  $k = 1, \dots, K$  do
6:       if  $k < K$  then
7:          $\text{VALUE} \leftarrow \bar{v}_n + \lambda_k c_n \mathcal{T}[n + 1, k + 1]$ 
8:          $\mathcal{T}[n, k] \leftarrow \max\{\text{VALUE}, \mathcal{T}[n + 1, k]\}$ 
9:       else
10:         $\mathcal{T}[n, k] \leftarrow \max\{\bar{v}_n, \mathcal{T}[n + 1, K]\}$ 
11:   return  $\mathcal{T}[1, 1]$ 

```

We end this subsection with the analysis of some economical and theoretical properties of the algorithm. From the economical standpoint, we note that, if the total orders used do not depend on the reported types of the agents, SORTED-ADS is maximal-in-its-range, leading thus to a truthful mechanism when paired with VCG-like payments. Furthermore, the resulting mechanism requires polynomial time both for the allocation and the payments. From the theoretical standpoint, we remark that it is possible to prove a $1/2$ approximation ratio for the SORTED-ADS algorithm in particular settings, e.g. when all the λ 's are constant (see the Appendix). The proof in the general case, though, is left here as an open conjecture.

Experimental evaluation

For a better comparison of the results, we adopted the same experimental setting used in (Gatti and Rocco 2013) and given to us by the authors. We briefly describe it, details can be found in the original paper. The experimental setting is based on *Yahoo! Webscope A3* dataset. Each bid is drawn from a truncated Gaussian distribution, where the mean and standard deviation are taken from the dataset, while quality is drawn from a beta distribution. The values of λ_s of the first 10 slots are $\{1.0, 0.71, 0.56, 0.53, 0.49, 0.47, 0.44, 0.44, 0.43, 0.43\}$. We considered two scenarios, one having $K = 5$ and one having $K = 10$. In both cases we let $N \in \{50, 60, \dots, 100\} \cup \{200, 300, \dots, 1000\}$. For each pair (K, N) , 20 instances were generated. We implemented our algorithms in the C++11 language and executed them on the OSX 10.10.3 operating system. The main memory was a 16GB 1600MHz DDR3 RAM, while the processor was an Intel Core i7-4850HQ CPU. We compiled the source with GNU g++ version 4.9.1. Parallelization was achieved using OpenMP version 4.0. Please refer to the Appendix for a more thorough analysis and for the complete boxplot version of the following plots.

DOMINATED-ADS algorithm

We study the average number of ads that survive DOMINATED-ADS in Figure 1. The upper bounding strategy needed in Lemma 14 was implemented using a $O(NK)$ time algorithm.

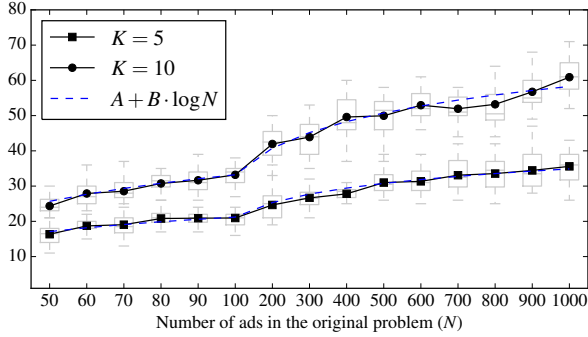


Figure 1: Average number of ads after running the DOMINATED-ADS algorithm.

Notice that the number of removed ads already considerable when N is small (e.g., $N \approx 50$), and that the prune ratio increases dramatically as N grows (for instance, the prune ratio is approximately 96% on average when $N = 1,000$ and $K = 5$). Experiments show that the number of surviving ads is of the form $\tilde{N} = A + B \cdot \log N$ for suitable values^{2,3} of A and B .

In Figure 2 we report the average running time for DOMINATED-ADS. The graph shows that the running time depends quadratically on the original problem size N , but it remains negligible (around 20ms) even when $N = 1,000$.

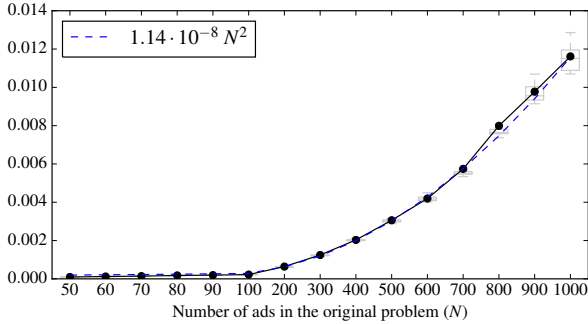


Figure 2: Avg. DOMINATED-ADS run. time [s]. This is independent of K and grows as $\approx 1.14 \cdot 10^{-8} N^2$ ($R^2 > 0.99$).

COLORED-ADS algorithm

Figure 3 shows the average running time for COLORED-ADS, implemented in its randomized version. The number of iterations was set to $e^K \log 2$, i.e. 50% probability of finding

²For $K = 5$, we have $A \approx -6.1$, $B \approx 5.9$ and $R^2 > 0.98$, where R^2 is the coefficient of determination of the fit.

³For $K = 10$, we have $A \approx -16.9$, $B \approx 10.9$ and $R^2 > 0.98$.

the optimal allocation. The algorithm is run on the reduced instances produced by DOMINATED-ADS. As a result of the shrinking process, we point out that the running times for $N = 50$ and $N = 1,000$ are comparable. We also remark that, in order for COLORED-ADS to be applicable in real-time contexts when $K = 10$, some sort of hardware scaling is necessary, as a running time of ≈ 0.5 seconds per instance is likely to be too expensive for many practical applications. When $K = 5$, though, no scaling is necessary, and the algorithm proves to be extremely efficient.

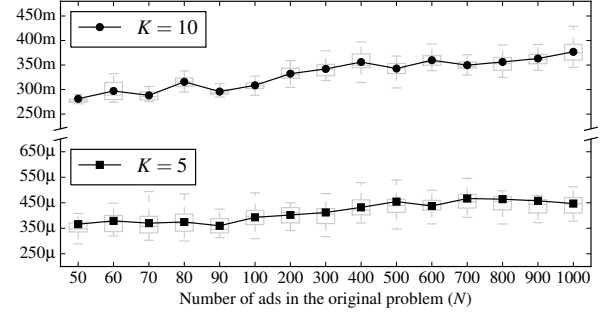


Figure 3: Avg. COLORED-ADS run. time [s] ($e^K \log 2$ runs).

SORTED-ADS algorithm

In figure 4 we study the approximation ratio of SORTED-ADS. Optimal allocation values were computed using an exponential branch-and-bound algorithm, similar to that of (Gatti and Rocco 2013). For each shrunk problem instance, we generated $2K^3$ random orders, and used SORTED-ADS to approximate the optimal allocation. Surprisingly, even though the number of iterations is relatively low, approximation ratios prove to be very high, with values $> 97\%$ in the worst case, with both medians and means always $> 99\%$. Finally, in Figure 5 we report the corresponding computation times. These prove to be in the order of a handful of milliseconds, thus making SORTED-ADS suitable in real-time contexts even for a large number of ads.

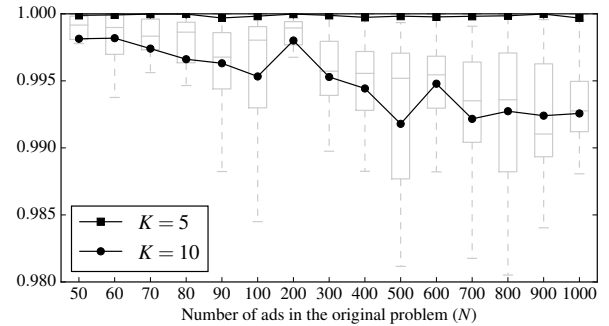


Figure 4: Avg. SORTED-ADS approx. ratio ($2K^3$ orders).

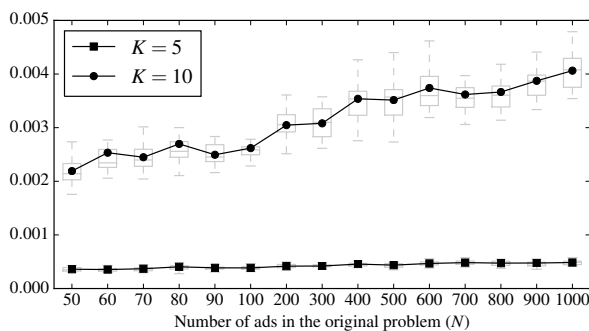


Figure 5: Avg. SORTED-ADS running time [s] ($2K^3$ orders).

Conclusions and future work

In this paper, we provide several contributions about the ad/position dependent cascade model for sponsored search auctions. This model is commonly considered the one providing the best tradeoff between accuracy and computational complexity, but no algorithm suitable for real-world applications is known. Initially, we study the inefficiency of GSP and VCG with the only position dependent cascade model w.r.t. the VCG with the ad/position dependent cascade model, analyzing PoA and PoS both of the social welfare and the auctioneer’s revenue. Our results show that the inefficiency is in most cases unbounded and that the dispersion over the equilibria is extremely large, suggesting that the GSP, as well as the VCG with position dependent cascade model, presents severe limits. Subsequently, we provide three algorithms and we experimentally evaluate them with a real-world dataset. Our first algorithm reduces the size of the instances, discarding ads. Empirically, the number of non-discarded ads is logarithmic. Our second algorithm finds the optimal allocation with high probability, requiring a short time (less than 1ms when $K = 5$ even with 1,000 ads), but too long for real-time applications when $K = 10$. Our third algorithm approximates the optimal allocation in very short time (less than 5ms even with 1,000 ads) providing very good approximations (>0.98). This shows that the ad/dependent cascade model can be effectively used in practice. Furthermore, we prove that our approximation algorithm has a theoretical bound of $1/2$ in restricted domains and we leave open the proof that the bound holds in general settings.

In future, we will extend our analysis and our algorithms to models in which there is a contextual graph among the ad, and in which there are multiple slates.

References

[Alon, Yuster, and Zwick 1994] Alon, N.; Yuster, R.; and Zwick, U. 1994. Color-coding: A new method for finding simple paths, cycles and other small subgraphs within large graphs. In *ACM STOC*, 326–335.

[Bentley 1979] Bentley, J. L. 1979. Decomposable searching problems. *Information Processing Letters* 8(5):244 – 251.

[Cooley and Tukey 1965] Cooley, J. W., and Tukey, J. W.

1965. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation* 19(90):297–301.

[Craswell et al. 2008] Craswell, N.; Zoeter, O.; Taylor, M.; and Ramsey, B. 2008. An experimental comparison of click position-bias models. In *WSDM*, 87–94.

[Evangelos and Telelis 2010] Evangelos, E., and Telelis, O. 2010. Discrete strategies in keyword auctions and their inefficiency for locally aware bidders. In *WINE*, 523–530.

[Fotakis, Krysta, and Telelis 2011] Fotakis, D.; Krysta, P.; and Telelis, O. 2011. Externalities among advertisers in sponsored search. In *SAGT*, 105–116.

[Gatti and Rocco 2013] Gatti, N., and Rocco, M. 2013. Which mechanism in sponsored search auctions with externalities? In *AAMAS*, 635–642.

[Gatti et al. 2015] Gatti, N.; Rocco, M.; Serafino, P.; and Ventre, C. 2015. Cascade model with contextual externalities and bounded user memory for sponsored search auctions. In *AAMAS*, 1831–1832.

[Gomes, Immorlica, and Markakis 2009] Gomes, R.; Immorlica, N.; and Markakis, E. 2009. Externalities in keyword auctions: An empirical and theoretical assessment. In *WINE*, 172–183.

[Hegeman 2010] Hegeman, J. 2010. Facebook’s ad auction. *Talk at Ad Auctions Workshop*.

[Hirschberg 1975] Hirschberg, D. S. 1975. A linear space algorithm for computing maximal common subsequences. *Communications of the ACM* 18(6):341–343.

[Hüffner, Wernicke, and Zichner 2008] Hüffner, F.; Wernicke, S.; and Zichner, T. 2008. Algorithm engineering for color-coding with applications to signaling pathway detection. *Algorithmica* 52(2):114–132.

[IAB 2015] IAB. 2015. IAB internet advertising revenue report. 2014 full year results.

[Joachims et al. 2007] Joachims, T.; Granka, L.; Pan, B.; Hembrooke, H.; Radlinski, F.; and Gay, G. 2007. Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search. *ACM Transactions on Information Systems* 25(2):7.

[Kempe and Mahdian 2008] Kempe, D., and Mahdian, M. 2008. A cascade model for externalities in sponsored search. In *WINE*, 585–596.

[Kuminov and Tennenholtz 2009] Kuminov, D., and Tennenholtz, M. 2009. User modeling in position auctions: re-considering the GSP and VCG mechanisms. In *AAMAS*, 273–280.

[Leme and E. Tardos 2010] Leme, R. P., and E. Tardos, E. 2010. Pure and Bayes-Nash price of anarchy for generalized second price auction. In *FOCS*, 735–744.

[Mehlhorn and Näher 1990] Mehlhorn, K., and Näher, S. 1990. Dynamic fractional cascading. *Algorithmica* 5(1-4):215–241.

[Naor, Schulman, and Srinivasan 1995] Naor, M.; Schulman, L. J.; and Srinivasan, A. 1995. Splitters and near-optimal derandomization. In *FOCS*, 182–191.

- [Narahari et al. 2009] Narahari, Y.; Garg, D.; Narayanam, R.; and Prakash, H. 2009. *Game Theoretic Problems in Network Economics and Mechanism Design Solutions*. Springer.
- [Nisan and Ronen 2000] Nisan, N., and Ronen, A. 2000. Computationally feasible VCG mechanisms. In *ACM EC*, 242–252.
- [Varian and Harris 2014] Varian, H. R., and Harris, C. 2014. The VCG auction in theory and practice. *American Economic Review* 104(5):442–445.

Appendix

Inefficiency of GSP and VCG with PDC

GSP analysis

Lemma 1. *The GSP is not IR, when users follow the APDC.*

Proof. Consider a problem with $N > K \geq 2$ having:

- $v_1 = \dots = v_N = 1$,
- $q_1 = \dots = q_N = 1$,
- $c_1 = \dots = c_N = 0$,
- $\lambda_1 = \dots = \lambda_{K-1} = 1$.

In the GSP model, there exists an optimal allocation mapping ad 2 to slot 2. Therefore, agent 2 pays 1, but actually gains no value since all the c 's are equal to 0. Hence, agent 2 has a negative payoff even though she is acting truthfully. \square

Lemma 2. *The PoA of the social welfare in the GSP when users follow APDC is unbounded.*

Proof. Consider a problem with $N = K = 2$ having:

- $v_1 = \varepsilon, v_2 = 1$,
- $q_1 = q_2 = 1$,
- $c_1 = \varepsilon, c_2 = 1 - \varepsilon$,
- $\lambda_1 = 1$,
- $\hat{v}_1 = 1/\varepsilon, \hat{v}_2 = \varepsilon^2/2$.

The utilities for the bidders are

$$u_1 = \bar{v}_1 - q_2 \hat{v}_2 = \varepsilon - \varepsilon^2/2,$$

$$u_2 = c_1 \lambda_1 \bar{v}_2 - 0 = \varepsilon.$$

Furthermore, the GSP allocates ad 1 before ad 2, with a social welfare of 2ε . The truthful strategy profile is a Nash equilibrium, as if bidder 1 made a bid $< \varepsilon/2$ her new utility would be $u'_1 = \varepsilon - \varepsilon^2 < u_1$; analogously, if bidder 2 made a bid > 1 , then her utility would be $u'_2 = 0 < u_2$. In the VCG with APDC model, the optimal allocation has a social welfare of $1 + \varepsilon(1 - \varepsilon)$. Letting $\varepsilon \rightarrow 0$, we obtain a ratio of $\frac{2\varepsilon}{1 + \varepsilon(1 - \varepsilon)} \rightarrow 0$. This concludes the proof. \square

Lemma 3. *The PoA of the social welfare in the GSP when users follow APDC is $\geq K$, in the restricted case the agents do not overbid.*

Proof. Consider a problem with $N > K \geq 2$ having:

- $v_1 = \dots = v_N = 1$,
- $q_1 = \dots = q_N = 1$,
- $c_1 = 0, c_2 = \dots = c_N = 1$,
- $\lambda_1 = \dots = \lambda_{K-1} = 1$,
- $\hat{v}_1 = \dots = \hat{v}_N = 1$.

The allocation mapping ad i to slot i for each $i = 1, \dots, K$ is optimal for the GSP, resulting in a social welfare of 1. Furthermore, the truthful strategy profile is a Nash equilibrium, since no agent would benefit from changing the reported type to a lower value. On the other hand, every optimal allocation for VCG with APDC allocates ad 1 in slot K , thus resulting in a social welfare of K . This concludes the proof. \square

Lemma 4. *The PoA of the revenue in the GSP when users follow APDC is unbounded (even without overbidding).*

Proof. Consider a problem with $N > K \geq 2$ having:

- $v_1 = 1, v_2 = \dots = v_N = 1 - \varepsilon$,
- $q_1 = \dots = q_N = 1$,
- $c_1 = 0, c_2 = \dots = c_N = 1$,
- $\lambda_1 = \dots = \lambda_{K-1} = 1$,
- $\hat{v}_1 = 1, \hat{v}_2 = \dots = \hat{v}_N = 0$.

The allocation mapping ad i to slot i for each $i = 1, \dots, K$ is optimal for the GSP. The revenue of the auctioneer is 0, each bidder been charged a payment of 0. The above strategy profile is a Nash equilibrium, as the first bidder obviously has no incentive in decreasing its reported type, while each other bidder can be allocated in the first slot only bidding a value larger than or equal to 1, but this would lead a payment $1 > 1 - \varepsilon$, generating thus a strictly negative utility. The VCG with APDC admits an optimal allocation featuring ad 1 in the last slot (slot K). In this case, bidder 1 is charged a payment $1 - \varepsilon$, leading the auctioneer to have a strictly positive revenue. This concludes the proof. \square

Lemma 5. *The PoS of the social welfare in the GSP when users follow APDC is 1 (even without overbidding).*

Proof. Since the SW is calculated in the same way, independently on the allocation function, the PoS of the SW is ≥ 1 . It is then sufficient to show that in at least one circumstance the allocation of GSP and VCG with APDC coincide. To this end, consider a problem with $N > K \geq 2$ having:

- $v_1 = \dots = v_N = 1$,
- $q_1 = \dots = q_N = 1$,
- $c_1 = \dots = c_N = 1$,
- $\lambda_1 = \dots = \lambda_{K-1} = 1$,
- $\hat{v}_1 = \dots = \hat{v}_N = 1$.

It is easy to see that the allocation mapping ad i to slot i for each $i = 1, \dots, K$ is optimal with both the mechanisms. Furthermore, it is a Nash equilibrium, since no agent benefits from misreporting her true type. This concludes the proof. \square

Lemma 6. *The PoS of the revenue in the GSP when users follow APDC is 0 (even without overbidding).*

Proof. Consider a problem with $N = K = 2$ having:

- $v_1 = 1, v_2 = 1/3$,
- $q_1 = q_2 = 1$,
- $c_1 = 1, c_2 = 1/2$,
- $\lambda_1 = 1$,
- $\hat{v}_1 = 1, \hat{v}_2 = 1/3$.

The GSP allocates ad 1 before ad 2, with a revenue of $1/3$. Notice that the above strategy profile constitutes a Nash equilibrium, as bidder 1 can only decrease its utility from $2/3$ to $1/2$ when ad 1 is allocated below ad 2, while bidder 2 can only decrease her utility from $1/3$ to $-2/3$ if ad 2 is allocated in the first slot. While the APDC allocates ad 1 before ad 2 as well, the revenue for the system is 0, since bidder 2 is not charged any payment (not being subject to competition for the second slot), and bidder 1 is charged $\bar{v}_2 - \lambda_1 c_1 \bar{v}_2 = 0$. This concludes the proof. \square

VCG with PDC analysis

Lemma 7. *The VCG with PDC is neither IR nor DSIC, when users follow APDC.*

Proof. Consider a problem with $N > K \geq 2$ having:

- $v_1 = \dots = v_N = 1$,
- $q_1 = \dots = q_N = 1$,
- $c_1 = \dots = c_N = 0$,
- $\lambda_1 = \dots = \lambda_K = 1$.

In the VCG with PDC, there exists an optimal allocation mapping ad 2 to slot 2. Therefore, agent 2 pays 1, but actually she gains no value since all the c 's are equal to 0. Hence, agent 2 has a negative payoff even though she bids truthfully and the mechanism is not IR nor DSIC. This concludes the proof. \square

Lemma 8. *The PoA of the social welfare in the VCG with PDC when users follow APDC is unbounded.*

Proof. Consider a problem with $N = K = 2$ having:

- $v_1 = 0, v_2 = 1$,
- $q_1 = q_2 = 1$,
- $c_1 = 0, c_2 = 1$,
- $\lambda_1 = 1/2$,
- $\hat{v}_1 = 4, \hat{v}_2 = 0$.

The utilities for the bidders are both 0 in the VCG with PDC model. The VCG with PDC allocates ad 1 before ad 2, as $q_1 \hat{v}_1 + \lambda_1 q_2 \hat{v}_2 > q_2 \hat{v}_2 + \lambda_1 q_1 \hat{v}_1$. The above strategy profile is a Nash equilibrium, as bidder 1 does not gain more by reducing her bid, while, if bidder 2 made any bid > 4 , her utility would be $\bar{v}_2 - (1 - \lambda)q_1 \hat{v}_1 < 0$. The social welfare produced by the VCG with PDC is 0. On the other hand, the VCG with APDC allocated ad 2 before ad 1, producing a strictly positive social welfare. This concludes the proof. \square

Lemma 9. *The PoA of the social welfare in the VCG with PDC when users follow APDC is $\geq K$, in the restricted case the agents do not overbid.*

Proof. The same setting used in the proof of Lemma 3 applies here. \square

Lemma 10. *The PoA of the revenue in the VCG with PDC when users follow APDC is unbounded (even without over-bidding).*

Proof. Consider a problem with $N = K \geq 2$ having:

- $v_1 = \dots = v_N = 1$,
- $q_1 = \dots = q_N = 1$,
- $c_1 = \dots = c_N = 1$,
- $\lambda_1 = \dots = \lambda_{K-1} = 1/2$,
- $\hat{v}_1 = \dots = \hat{v}_N = 1$.

Any allocation is optimal for the VCG with PDC, charging each bidder a payment of zero. Furthermore, the above strategy profile is a Nash equilibrium for the VCG with PDC, since in any allocation the utility of each bidder is the same. Thus, the revenue of the auctioneer with VCG with PDC is 0. On the other hand, in the VCG with APDC, all the bidders except the one allocated in slot K are charged a strictly positive payment. This concludes the proof. \square

Lemma 11. *The PoS of the social welfare in the VCG with PDC when users follow APDC is 1.*

Proof. The same setting used in the proof of Lemma 5 applies here. \square

Lemma 12. *The PoS of the revenue in the VCG with PDC when users follow APDC is 0.*

Proof. Consider a problem with $N = K = 2$ having:

- $v_1 = 1, v_2 = 0$,
- $q_1 = q_2 = 1$,
- $c_1 = 1, c_2 = 1/2$,
- $\lambda_1 = 1/2$,
- $\hat{v}_1 = 1, \hat{v}_2 = 1/2$.

Both VCG with PDC and VCG with APDC allocate ad 1 before ad 2. With VCG with PDC, bidder 1 has a utility of $3/4$, while bidder 2 has a utility of 0. Notice that the above strategy profile is a Nash equilibrium of the VCG with PDC: if bidder 1 made a bid $< 1/2$, her utility would be $1/4$; if bidder 2 made a bid > 1 , her utility would be $-1/2$. The revenue for VCG with PDC is $1/4$, while that of VCG with APDC is 0. This concludes the proof. \square

Lemma 13. *The PoS of the revenue in the VCG with PDC when users follow APDC is ≤ 1 , in the restricted case the agents do not overbid.*

Proof. Consider a problem with $N > K \geq 2$ having:

- $v_1 = \dots = v_N = 1$,
- $q_1 = \dots = q_N = 1$,
- $c_1 = \dots = c_N = 1$,
- $\lambda_1 = \dots = \lambda_{K-1} = 1/2$,
- $\hat{v}_1 = \dots = \hat{v}_N = 1$.

Notice that the strategy is a Nash equilibrium of the VCG with PDC, as no bidder can overbid, and no bidder benefits from lowering her reported type. Furthermore, since all the ad continuation probabilities are equal to 1, the revenues of the two systems are equal, thus proving the lemma. This concludes the proof. \square

DOMINATED-ADS algorithm

Lemma 14. *Let $\lambda_{\max} = \max\{\lambda_s : s \in \mathcal{K}\}$, and B an upper bound⁴ of the value of the optimal allocation; also, let $\mathcal{D} = [0, \lambda_{\max}] \times [0, B]$. Given two ads a, b with parameters (\bar{v}_a, c_a) and (\bar{v}_b, c_b) , consider the affine function $w_{a,b} : \mathcal{D} \rightarrow \mathbb{R}$, defined as*

$$w_{a,b}(x, y) = \det \begin{pmatrix} x & -y & 1 \\ 1 & \bar{v}_b & c_b \\ 1 & \bar{v}_a & c_a \end{pmatrix}$$

If the minimum of $w_{a,b}$ over \mathcal{D} is greater than 0, then a dominates b .

⁴ More precisely, it is enough that B is a (weak) upper bound of the quantity

$$\tilde{B} = \max_i \{\lambda_i \cdot \text{ALLOC}(i+1, K)\},$$

where $\text{ALLOC}(i, K)$ is the value of the optimal allocation of the problem having as slots the set $\mathcal{K}_i = \{i, \dots, K\} \subseteq \mathcal{K}$.

Proof. Without loss of generality, we introduce a fictitious slot, say $K+1$, with $\lambda_K = 0$, representing the case in which an ad is not displayed. By simple calculations, it can be observed that, given an allocation θ and two allocated ads a, b such that b is allocated above a (and a may be allocated in slot $K+1$), swapping a and b leads to an improvement of the social welfare when the following condition holds:

$$\bar{v}_a + c_a D + \bar{v}_b c_a E > \bar{v}_b + c_b D + \bar{v}_a c_b E$$

where:

$$\bullet D = \frac{\sum_{s \in \{\text{slot}_\theta(b), \dots, \text{slot}_\theta(a)\}} q_{\text{ad}_\theta(s)} v_{\text{ad}_\theta(s)} C_\theta(\text{ad}_\theta(s)) \Lambda_s}{C_\theta(b) \Lambda_{\text{slot}_\theta(b)}},$$

$$\bullet E = \frac{C_\theta(a) \Lambda_{\text{slot}_\theta(a)}}{C_\theta(b) \Lambda_{\text{slot}_\theta(b)}}.$$

Notice that both D and E do not depend on the parameters (v, q, c) of a and b .

Ad a will be always swapped with ad b independently of θ (i.e., independently of the slots in which they are allocated and of how the other ads are allocated) and therefore ad a dominates ad b , if the above inequality holds for every feasible value of D and E . More precisely, D is lower bounded by 0 and upper bounded by the value of the best allocation (i.e., B), while E is lower bounded by 0 and upper bounded by λ_{\max} . Rearranging the above inequality and replacing D with y and E with x , we obtain:

$$x(\bar{v}_b c_a - \bar{v}_a c_b) + y(c_a - c_b) + \bar{v}_a - \bar{v}_b = w_{a,b}(x, y).$$

This concludes the proof. \square

CONST- λ algorithm

The idea behind the CONST- λ algorithm is that an upper bound for $\text{ALLOC}(k, K)$ is the optimal allocation of the problem where all the λ values are replaced with λ_{\max} . This value can be (exactly) computed in $O(NK)$ time using SORTED-ADS, by sorting on decreasing $\bar{v}/(1 - \lambda_{\max} c)$

DECOUPLE algorithm

The idea behind the DECOUPLE algorithm is that

$$\begin{aligned} \text{ALLOC}(k, K) &\geq \text{UB}_{\text{decouple}}(k) \\ &= \bar{v}_1^* + \bar{v}_2^* c_1^* \lambda_k + \dots \\ &\quad + \bar{v}_{K-k+1}^* (c_1^* \dots c_{K-k}^*) (\lambda_k \dots \lambda_{K-1}) \end{aligned}$$

where the \bar{v}^* values represent the \bar{v} values of the ads, ordered in decreasing order (i.e. \bar{v}_1^* is the maximum among all the \bar{v} values); same for the c^* values. Notice that we can easily compute $\bar{v}_1^*, \dots, \bar{v}_K^*$ in $O(K \log K + N)$ time using a linear selection algorithm (e.g. introselect) and then sorting the topmost K values with an efficient sort algorithm (the same holds for the c^* values).

In order to efficiently compute \tilde{B} , we need to compute the values $\text{UB}_{\text{decouple}}(1), \dots, \text{UB}_{\text{decouple}}(K)$ in time sub-quadratic in K , as in the worst case $K = N$. We define

$$a_k = \bar{v}_k^* (c_1^* \dots c_{k-1}^*), \quad b_k = (\lambda_1 \dots \lambda_{k-1}).$$

In particular, $a_1 = \bar{v}_1^*$ and $b_1 = 1$. Notice that we can assume without loss of generality that $b_k \neq 0$, $k = 1, \dots, K$. We also let

$$c_k = b_k \text{UB}_{\text{decouple}}(k).$$

It is easy to prove that the vector $\mathbf{c} = (c_K, \dots, c_1)$ is the (discrete) convolution of vectors $\mathbf{a} = (a_1, \dots, a_K)$ and $\mathbf{b} = (b_K, \dots, b_1)$. Indeed, for every $k \in \{1, \dots, K\}$:

$$\begin{aligned} (\mathbf{a} * \mathbf{b})_k &= \sum_{i=1}^k \mathbf{a}_i \mathbf{b}_{k-i+1} = \sum_{i=1}^k a_i b_{K-k+i} \\ &= \sum_{i=1}^k \bar{v}_i^* (c_1^* \dots c_{i-1}^*) (\lambda_1 \dots \lambda_{K-k+i-1}) \\ &= b_{K-k+1} \text{UB}_{\text{decouple}}(K-k+1) \\ &= c_{K-k+1} \\ &= \mathbf{c}_k. \end{aligned}$$

This means that we can determine \mathbf{c} in $O(K \log K)$ time using the FFT algorithm (Cooley and Tukey 1965). Once we have \mathbf{c} , it's trivial to compute $\text{UB}_{\text{decouple}}(k)$ for $k = 1, \dots, K$ in $O(K)$ time. The final complexity for the algorithm is then $O(K \log K + N)$.

Fast dominance computation

In order to show how to compute $|\text{dom}(a)|$ for each ad a , we begin with this simple lemma:

Lemma 19. *For any fixed $(x, y) \in \mathcal{D}$, the order $\prec_{(x,y)}$ over the set of ads, defined as*

$$a \prec_{(x,y)} b \iff w_{a,b}(x, y) > 0,$$

*is a total order*⁵.

Proof. We have, by the properties of the determinant:

$$\begin{aligned} w_{a,b}(x, y) &= \det \begin{pmatrix} x & -y & 1 \\ 1 & \bar{v}_b & c_b \\ 1 & \bar{v}_a & c_a \end{pmatrix} \\ &= \det \begin{pmatrix} x & -y & 1 \\ 1 - c_b x & \bar{v}_b + c_b y & 0 \\ 1 - c_a x & \bar{v}_a + c_a y & 0 \end{pmatrix} \\ &= (1 - c_b x)(\bar{v}_a + c_a y) - (1 - c_a x)(\bar{v}_b + c_b y) \end{aligned}$$

Hence, $w_{a,b}(x, y) > 0$ if, and only if,

$$\frac{1 - c_b x}{\bar{v}_b + c_b y} > \frac{1 - c_a x}{\bar{v}_a + c_a y}$$

As we assumed x and y to be fixed parameters, we see that $\prec_{(x,y)}$ defines a total order over the set of ads. \square

Lemma 19 and Lemma 15 together prove that the domination partial order, \prec , is the intersection of the four total orders $\prec_{(0,0)}$, $\prec_{(0,B)}$, $\prec_{(\lambda_{\max},0)}$ and $\prec_{(\lambda_{\max},B)}$. This immediately suggests the idea of working with ranks; in particular, we define $\rho_{(x,y)}(a)$ as the rank of ad a in the total order $\prec_{(x,y)}$, and introduce the 4-dimensional vector

$$\mathbf{rk}(a) = \begin{pmatrix} \rho_{(0,0)}(a) \\ \rho_{(0,B)}(a) \\ \rho_{(\lambda_{\max},0)}(a) \\ \rho_{(\lambda_{\max},B)}(a) \end{pmatrix}.$$

⁵We assume a deterministic tie-breaking condition is used to handle the case when $w_{a,b}(x, y) = 0$. Hence, from now on we will assume without loss of generality that $w_{a,b}(x, y) \neq 0$ for all $(x, y) \in \mathcal{D}$ and $a, b \in \mathcal{A}$

The condition of Lemma 15 is then equivalent to the following:

Lemma 20. *Given two ads a and b , $\mathbf{rk}(a) < \mathbf{rk}(b)$ iff $a \prec b$.*

Here $<$ denotes component-wise comparison. Using Lemma 20, we see that

$$\text{dom}(a) = \{b \in \mathcal{A} : \mathbf{rk}(b) < \mathbf{rk}(a)\}$$

In other words, we can cast the problem of counting the number of ads dominating a to a geometric dominance counting problem in four dimensions. This problem is very well-studied in the literature, with many results showing algorithms computing dominance counts for every point in the point set in $O(N \text{ poly log } N)$ time. For instance, we could use a 4-dimensional range tree (Bentley 1979) to calculate $|\text{dom}(a)|$ for every ad a in batch in $O(N \log^4 N)$ time.

To lower the exponent of the log factor, we exploit the structure of the problem a bit further. We define the two sets:

$$\begin{aligned} \text{dom}^-(a) &= \{b \in \mathcal{A} : b \prec a \wedge c_b \leq c_a\} \\ \text{dom}^+(a) &= \{b \in \mathcal{A} : b \prec a \wedge c_b > c_a\} \end{aligned}$$

It is trivial to verify that $\text{dom}^-(a) \cup \text{dom}^+(a) = \text{dom}(a)$ and $\text{dom}^-(a) \cap \text{dom}^+(a) = \emptyset$, i.e. that $\text{dom}^-(a)$ and $\text{dom}^+(a)$ are a partition of $\text{dom}(a)$. Hence,

$$|\text{dom}(a)| = |\text{dom}^-(a)| + |\text{dom}^+(a)|.$$

Notice that

$$\frac{\partial}{\partial y} w_{a,b}(x, y) = c_a - c_b$$

This means that when $c_b \leq c_a$, $w_{a,b}(x, y) \geq w_{a,b}(x, 0)$, for all $(x, y) \in \mathcal{D}$. Analogously, when $c_b > c_a$, we have $w_{a,b}(x, y) \geq w_{a,b}(x, B)$ for all $(x, y) \in \mathcal{D}$. Therefore, it makes sense to introduce the following restricted ranks:

$$\mathbf{rk}^-(a) = \begin{pmatrix} \rho_{(0,0)}(a) \\ \rho_{(\lambda_{\max},0)}(a) \end{pmatrix}, \quad \mathbf{rk}^+(a) = \begin{pmatrix} \rho_{(0,B)}(a) \\ \rho_{(\lambda_{\max},B)}(a) \end{pmatrix}$$

We formalize the observation above in the next Lemma:

Lemma 21. *Given two ads a and b :*

- $b \in \text{dom}^-(a) \iff \mathbf{rk}^-(b) < \mathbf{rk}^-(a)$;
- $b \in \text{dom}^+(a) \iff \mathbf{rk}^+(b) < \mathbf{rk}^+(a)$.

Assume we have a data structure RT supporting each of the following operations in $O(\log N)$ time:

- $\text{RT}.\text{INSERT}(\mathbf{v})$ insert the 2-dimensional point \mathbf{v} in the structure; point \mathbf{v} is not already present.
- $\text{RT}.\text{ERASE}(\mathbf{v})$ deletes the 2-dimensional point \mathbf{v} from the structure; point \mathbf{v} is guaranteed to be present. Furthermore, after the first call to $\text{SC}.\text{ERASE}$ is made, the structure no longer accepts calls to $\text{DS}.\text{INSERT}$.
- $\text{RT}.\text{DOMINANCE}(\mathbf{v})$ counts the number of points \mathbf{w} in the structure, satisfying $\mathbf{w} < \mathbf{v}$.

Such a data structure exists and is well-studied in the literature. It basically consists of a 2-dimensional range tree combined with the technique of *dynamic fractional cascading* (Mehlhorn and Näher 1990).

The RT data structure can be used to compute $|\text{dom}(a)|$ for every ad a in batch in $O(N \log N)$ time, as streamlined in Algorithm 4.

Algorithm 4 FAST-DOMINATED-ADS

```

1: procedure FAST-DOMINATED-ADS(ads, slots)
2:   Determine  $\mathcal{D}$ , as defined in Lemma 14
3:    $\text{RT}^+ \leftarrow \text{new RT}$ 
4:    $\text{RT}^- \leftarrow \text{new RT}$ 
5:   for each ad  $a \in \mathcal{A}$  do
6:      $\text{RT}^-.\text{INSERT}(\mathbf{rk}^-(a))$ 
7:   for each ad  $a \in \mathcal{A}$ , sorted according to decreasing  $c$ , do
8:      $|\text{dom}^+(a)| \leftarrow \text{RT}^+.\text{DOMINANCE}(\mathbf{rk}^+(a))$ 
9:      $|\text{dom}^-(a)| \leftarrow \text{RT}^-.\text{DOMINANCE}(\mathbf{rk}^-(a))$ 
10:     $|\text{dom}(a)| \leftarrow |\text{dom}^+(a)| + |\text{dom}^-(a)|$ 
11:     $\text{RT}^+.\text{INSERT}(\mathbf{rk}^+(a))$ 
12:     $\text{RT}^-.\text{ERASE}(\mathbf{rk}^-(a))$ 
13:   Discard all ads  $a$  having  $|\text{dom}(a)| > K$ 

```

COLORED-ADS algorithm

Lemma 18. *In at least one optimal allocation all the available slots are allocated.*

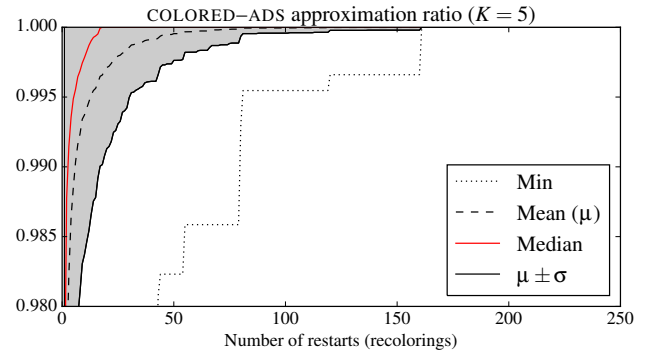
Proof. Let θ be any optimal allocation for the problem at hand. If it respects the condition of the lemma, the lemma is trivially true. Otherwise, we show that it is possible to build an optimal allocation θ^* from θ , respecting the condition.

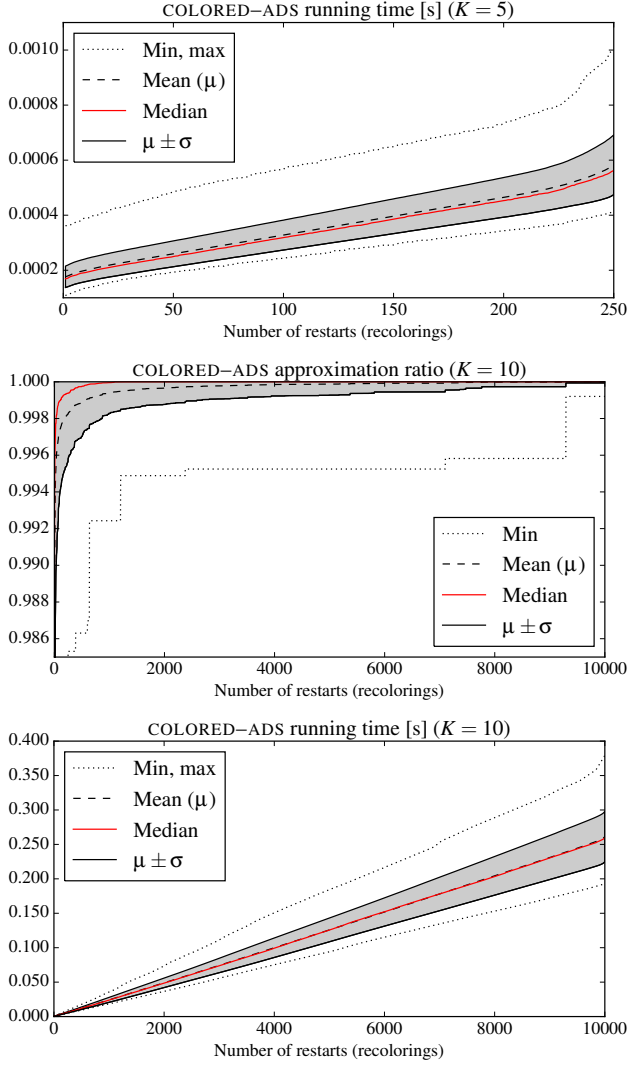
Let k^* the maximum index of any empty slot of θ . If it is the last slot, take any non allocated ad, and allocate it to k^* . Such an ad must exist because we always assume $N \geq K$. As the value of the new allocation is not less than that of θ , we conclude that the new allocation is optimal.

If, on the other hand, $k^* \neq K$, we act as follows: we move all the ads allocated in slots $k^* + 1$ to K one slot ahead. This operation does not decrease the value of the allocation, and at the same time leaves slot K empty. We then operate as in the previous case, filling slot K with any non allocated ad.

We repeat this algorithm until no slot is left vacant. \square

Time-approximation tradeoffs





SORTED-ADS algorithm

Space-efficient algorithm

We modify the idea in (Hirschberg 1975) to find a $O(NK)$ time, $O(N)$ memory implementation of SORTED-ADS. In particular, we need to show that there exists a $O(NK)$ time and $O(N)$ memory algorithm able to find the least slot k^* , respecting Definition 7.

Definition 7. Slot k is a split-point if there exists an optimal allocation such that slots $1, \dots, k-1$ are occupied only by ads $\leq N/2$ and slots k, \dots, K are occupied only by ads $> N/2$.

Conceptually, we scan all the possible values of k , one by one from 1 up until K until we find the first split-point. Once a tentative value for k^* has been chosen, it is enough to test whether the best allocation satisfying the condition in Definition 7 is an optimal allocation for the original problem. The split-point k breaks the optimal allocation fulfilling the condition into two natural pieces: the first using only ads in $\{1, \dots, N/2\}$ and slots $1, \dots, k-1$, and the second

piece using ads in $\{N/2 + 1, \dots, N\}$ and slots k, \dots, K . We let the social welfare of the first piece be W_1 , and that of the second piece be W_2 ; also, we let C_1 be the product of the continuation probabilities of the ads in the first piece. The value of the optimal allocation satisfying the split-point condition is therefore

$$SW = W_1 + \Lambda_{k-1} C_1 W_2$$

Since by definition the allocation is optimal, the second piece of the allocation must be itself an optimal allocation for the problem having $\mathcal{A} = \{N/2 + 1, \dots, N\}$, $\mathcal{K} = \{k, \dots, K\}$. Notice that we already know how to find W_2 , for all the possible values of k , in $O(NK)$ time and $O(N)$ memory. We can therefore treat W_2 as a known constant and focus on the problem of determining the allocation of the first piece maximizing SW. The crucial point is that we can find the maximum of W_1 for each k in dynamic programming. Indeed, the objective function is the same we would have in the allocation problem in which we have to allocate ads $\{1, \dots, N/2\} \cup \{X\}$ in slots $1, \dots, k$ under the constraints:

- $\bar{v}_X = W_2$ (notice that W_2 is known, while the values c_X and λ_k can be disregarded not affecting the objective function).
- Ad X must be allocated in slot k .

$\frac{1}{2}$ -approximation ratio

Definition 8. Given a total order \prec_{ads} , we define $\Theta_{\prec_{ads}}$ as the set of allocations θ where:

- for every $a, a' \in \theta$, $\text{slot}_\theta(a) < \text{slot}_\theta(a')$ only if $a \prec_{ads} a'$ (notice that the constraint is only over the allocated ads, each non-allocated ad may appear everywhere in \prec_{ads});
- there are no empty slots before an allocated slot.

Given a \prec_{ads} , the set $\Theta_{\prec_{ads}}$ constitutes the range of our allocation algorithm.

Definition 9. We define $\theta_{\prec_{ads}}^*$ as the optimal allocation (maximizing the social welfare) among all the allocations in $\Theta_{\prec_{ads}}$.

Definition 10. We define θ^* as the optimal allocation (maximizing the social welfare) among all the allocations in Θ and \prec_{ads}^* as $\prec_{ads}^* = \{\prec_{ads} : \theta_{\prec_{ads}}^* = \theta^*\}$.

Definition 11. We define \prec_{ads}^N as: for every $a, a' \in A$, $a \prec_{ads}^N a'$ if and only if $\frac{\bar{v}_a}{1-c_a} \geq \frac{\bar{v}_{a'}}{1-c_{a'}}$ (ties are broken lexicographically).

When $\lambda_k = 1$ for every k , we have $\prec_{ads}^N \in \prec_{ads}^*$ for every value of the parameters as shown in (Kempe and Mahdian 2008).

Definition 12. We define \prec_{ads}^R as the reverse of \prec_{ads}^N .

We introduce a lemma that we use in our main result.

Lemma 22. When $\lambda_k = 1$ for every k , once A is restricted to the ads appearing in θ^* and the ads are labeled such that a is the a -th ad in the allocation, then for every a we have:

$$\frac{\bar{v}_a}{1-c_a} \geq \sum_{b=a}^K \bar{v}_b \prod_{h=i}^{b-1} c_h.$$

Proof. The proof is by mathematical induction.

Basis of the induction. The basis of the induction is for slots k and $k - 1$. We need to prove that:

$$\frac{\bar{v}_{k-1}}{1 - c_{k-1}} \geq \bar{v}_{k-1} + c_{k-1} \bar{v}_k. \quad (1)$$

Equation (1) can be rewritten as:

$$\frac{\bar{v}_{k-1}}{1 - c_{k-1}} \geq \bar{v}_k,$$

that holds since

$$\frac{\bar{v}_{k-1}}{1 - c_{k-1}} \geq \frac{\bar{v}_k}{1 - c_k}$$

(by definition of \prec_{ads}^N) and

$$\frac{\bar{v}_k}{1 - c_k} \geq \bar{v}_k$$

(by $c_k \in [0, 1]$).

Inductive step. We assume that following inequality holds

$$\frac{\bar{v}_{i+1}}{1 - c_{i+1}} \geq \sum_{j=i+1}^k \bar{v}_j \prod_{h=i+1}^{j-1} c_h,$$

and we want to prove

$$\frac{\bar{v}_{a_i}}{1 - c_{a_i}} \geq \sum_{j=i}^k \bar{v}_{a_j} \prod_{h=i}^{j-1} c_{a_h}. \quad (2)$$

Equation (2) can be written as:

$$\frac{\bar{v}_i}{1 - c_i} \geq \sum_{j=i+1}^k \bar{v}_j \prod_{h=i+1}^{j-1} c_h,$$

that holds since

$$\frac{\bar{v}_i}{1 - c_i} \geq \frac{\bar{v}_{i+1}}{1 - c_{i+1}}$$

(by definition of \prec_{ads}^N) and

$$\frac{\bar{v}_{i+1}}{1 - c_{i+1}} \geq \sum_{j=i+1}^k \bar{v}_j \prod_{h=i+1}^{j-1} c_h$$

(by assumption of the inductive step). This completes the proof. \square

We use the above lemma to derive the bound.

Theorem 1. When $\lambda_k = 1$ for every k , we have

$$r = \frac{SW(\theta_{\prec_{\text{ads}}}^*)}{SW(\theta^*)} \geq \frac{1}{2}$$

for every \prec_{ads} .

Proof. For the sake of presentation, we split the proof in several parts.

Ads ranking. Among all the possible \prec_{ads} , we can safely focus only on \prec_{ads}^R given that this order is the worst for our problem. This property follows from the results discussed in (Kempe and Mahdian 2008). More precisely, the authors show that, given any allocation θ , for every a, a' if $\text{slot}_\theta(a) < \text{slot}_\theta(a')$ and $\frac{\bar{v}_a}{1 - c_a} > \frac{\bar{v}_{a'}}{1 - c_{a'}}$ then switching a with a' never reduces the value of the allocation. Therefore, the social welfare with \prec_{ads}^R is the minimum w.r.t. that one with all the \prec_{ads} . (Interestingly, when $\prec_{\text{ads}} \not\prec \prec_{\text{ads}}^*$, $\theta_{\prec_{\text{ads}}}^*$ may prescribe that the number of allocated ads is strictly smaller than the number of ads allocated in θ^* .)

Allocations set. Given \prec_{ads}^R , we use a lower bound over the value $\theta_{\prec_{\text{ads}}^R}^*$ focusing only on a specific subset of allocations defined below. This subset may not contain the optimal allocation $\theta_{\prec_{\text{ads}}^R}^*$ and therefore we obtain a lower bound for r that may be not tight. However, in the following, we complete the result showing that the bound is tight. The allocations we focus on are of the form: $\langle z, z - 1, \dots, 1 \rangle$ with $z \in \{1, \dots, K\}$ where the ads are indexed on the basis on \prec_{ads}^N (i.e., $1 \prec_{\text{ads}}^N 2 \prec_{\text{ads}}^N \dots$). We study the value of r for each possible value of z under the constraints that, given z , it holds:

$$\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right) \geq \sum_{i=1}^{z+1} \left(\bar{v}_i \prod_{j=i+1}^{z+1} c_j \right) \quad (3)$$

$$\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right) \geq \sum_{i=1}^h \left(\bar{v}_i \prod_{j=i+1}^h c_j \right) \quad \forall h : 1 \leq h < z \quad (4)$$

with the exception that when $z = k$, only Constraints (4) are required. We notice that, for every auction instance, there always exists a value of z such that the above constraints hold. This value can be found by starting from $z = 1$ and increasing it until the constraints hold.

Ratio. Once z is fixed, a lower bound over r can be obtained by using the following upper bound over the value of θ^* : we compute θ^* as if $k = n$. Thus, the ratio r is:

$$r \geq \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\sum_{i=1}^n \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right)}.$$

Bound with $z = 1$. In this case, we have:

$$z = 1$$

$$\bar{v}_1 \geq \bar{v}_2 + c_2 \bar{v}_1 \quad (5)$$

$$r \geq \frac{\bar{v}_1}{\sum_{i=1}^n \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right)} \quad (6)$$

by applying Lemma 22, we can write r as:

$$r \geq \frac{\bar{v}_1}{\sum_{i=1}^n \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right)} \geq \frac{\bar{v}_1}{\bar{v}_1 + c_1 \frac{\bar{v}_2}{1 - c_2}},$$

by using Equation (5) we have:

$$r \geq \frac{\bar{v}_1}{\bar{v}_1 + c_1 \frac{\bar{v}_2}{1 - c_2}} \geq \frac{\bar{v}_1}{\bar{v}_1 + c_1 \bar{v}_1}.$$

The minimization of $\frac{\bar{v}_1}{\bar{v}_1 + c_1 \bar{v}_1}$ can be achieved by maximizing c_1 (i.e., when $c_1 = 1$), obtaining $r \geq \frac{1}{2}$.

Bound with $k > z > 1$. In this case, we can write r as:

$$\begin{aligned} r &\geq \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\sum_{i=1}^n \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right)} \\ &= \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^z c_j \sum_{i=z+1}^n \left(\bar{v}_i \prod_{j=z+1}^{i-1} c_j \right)}, \end{aligned}$$

by applying Lemma 22, we have:

$$\begin{aligned} r &\geq \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^z c_j \sum_{i=z+1}^n \left(\bar{v}_i \prod_{j=z+1}^{i-1} c_j \right)} \\ &\geq \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^z c_j \frac{\bar{v}_{z+1}}{1 - c_{z+1}}}, \end{aligned}$$

by using Equation (3) we have:

$$\begin{aligned} r &\geq \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^z c_j \frac{\bar{v}_{z+1}}{1 - c_{z+1}}} \\ &\geq \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^z c_j \sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}. \end{aligned}$$

We derive r w.r.t. \bar{v}_z and we show that the derivative is al-

ways positive:

$$\begin{aligned} \frac{\partial r}{\partial \bar{v}_z} &= \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) - \prod_{j=1}^{z-1} c_j \sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right)}{\left(\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^z c_j \sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right) \right)^2} \\ &= \frac{\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \left(1 - \prod_{j=i}^{z-1} c_j \prod_{j=i+1}^z c_j \right) \right)}{\left(\sum_{i=1}^z \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^z c_j \sum_{i=1}^z \left(\bar{v}_i \prod_{j=i+1}^z c_j \right) \right)^2} \\ &\geq 0, \end{aligned}$$

given that $\left(1 - \prod_{j=i}^{z-1} c_j \prod_{j=i+1}^z c_j \right) \geq 0$. Thus, we minimize r by minimizing \bar{v}_z . Due to Constraint (4) with $h = z - 1$, the minimum feasible value of \bar{v}_z is:

$$\bar{v}_z = (1 - c_z) \left(\sum_{i=1}^{z-1} \left(\bar{v}_i \prod_{j=i+1}^{z-1} c_j \right) \right),$$

substituting \bar{v}_z in r we have:

$$r \geq \frac{\sum_{i=1}^{z-1} \left(\bar{v}_i \prod_{j=i+1}^{z-1} c_j \right)}{\sum_{i=1}^{z-1} \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) + \prod_{j=1}^{z-1} c_j \sum_{i=1}^{z-1} \left(\bar{v}_i \prod_{j=i+1}^{z-1} c_j \right)}.$$

By applying iteratively the above steps for \bar{v}_h with h from $h = z - 2$ to $h = 2$, we obtain the same bound of the case with $z = 1$:

$$r \geq \frac{\bar{v}_1}{\bar{v}_1 + c_1 \bar{v}_1} \geq \frac{1}{2}.$$

Bound with $z = k$. Differently from the previous cases, here we do not use an upper bound over θ^* , but the exact value, and for the lower bound over θ_{ads}^{*R} we consider only the ads appearing in θ^* . In this case, we have:

$$r \geq \frac{\sum_{i=1}^k \left(\bar{v}_i \prod_{j=i+1}^k c_j \right)}{\sum_{i=1}^k \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right)},$$

we derive r w.r.t. \bar{v}_k , obtaining:

$$\begin{aligned} \frac{\partial r}{\partial \bar{v}_k} &= \frac{\sum_{i=1}^k \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \left(1 - \prod_{j=i}^{k-1} c_j \prod_{j=i+1}^k c_j \right) \right)}{\left(\sum_{i=1}^k \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \right) \right)^2} \\ &\geq 0, \end{aligned}$$

we minimize r by minimizing \bar{v}_k as:

$$\bar{v}_k = (1 - c_k) \left(\sum_{i=1}^{k-1} \left(\bar{v}_i \prod_{j=i+1}^{k-1} c_j \right) \right),$$

obtaining:

$$r \geq \frac{\sum_{i=1}^{k-1} \left(\bar{v}_i \prod_{j=i+1}^{k-1} c_j \right)}{\sum_{i=1}^{k-1} \left(\bar{v}_i \prod_{j=1}^{i-1} c_j \left(1 + (1 - c_k) c_i \prod_{j=i+1}^{k-1} c_j^2 \right) \right)}.$$

By applying iteratively the above steps for \bar{v}_h with h from $h = k - 2$ to $h = 2$ —the derivative is always positive—, we obtain:

$$r \geq \frac{\bar{v}_1}{\bar{v}_1 + \bar{v}_1 c_1 \left(1 - \prod_{i=2}^k c_j \right)} = \frac{1}{1 + c_1 \left(1 - \prod_{i=2}^k c_j \right)},$$

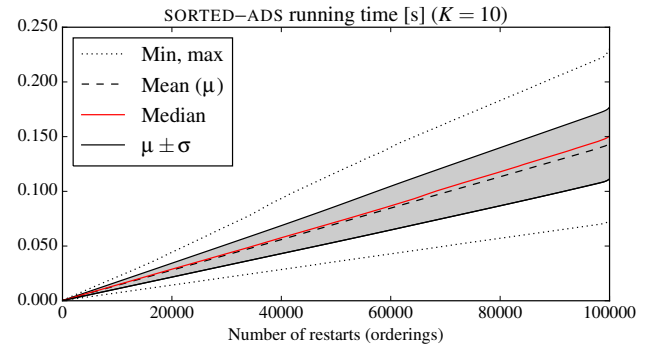
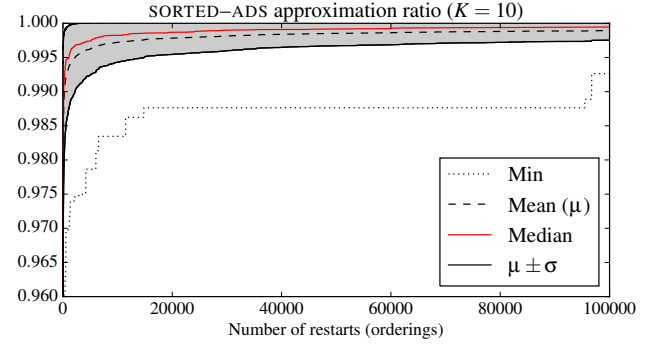
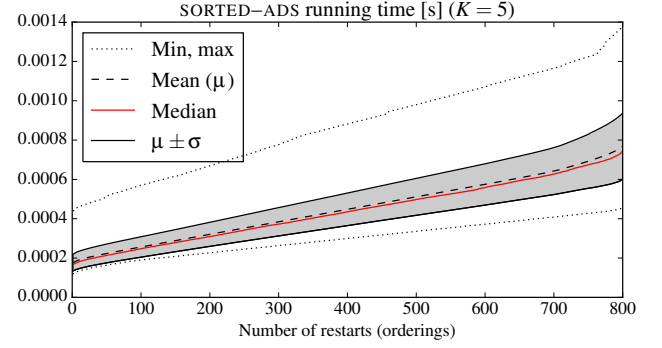
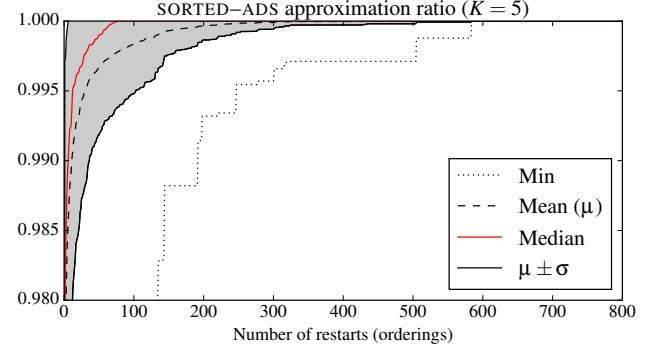
r is minimized when $c_1 = 1$ and $\prod_{i=2}^k c_j = 0$, obtaining

$$r \geq \frac{1}{2}.$$

□

The extension of the above result to the case in which $\lambda_k = \lambda$ for every k is straightforward. Indeed, it is sufficient to observe that in this case we can redefine the continuation probabilities as $c'_i = c_i \lambda$ for every i and subsequently redefine $\lambda_k = 1$ for every k .

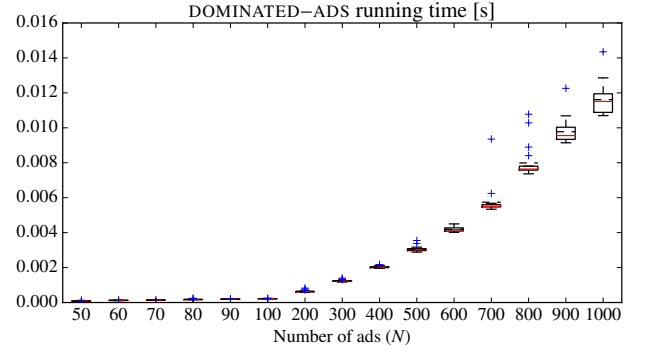
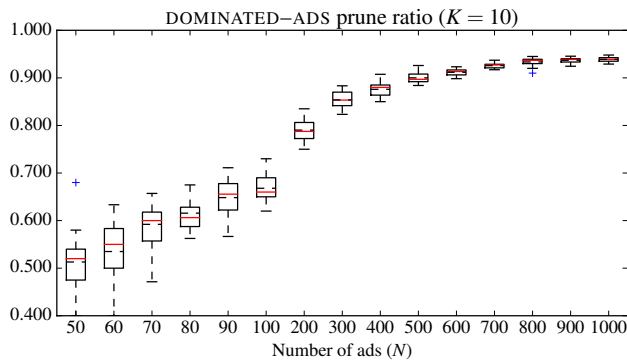
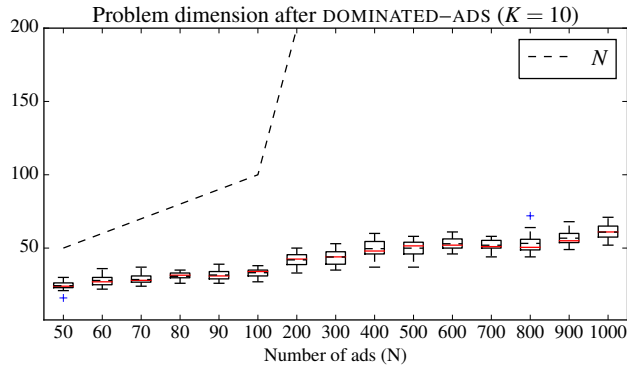
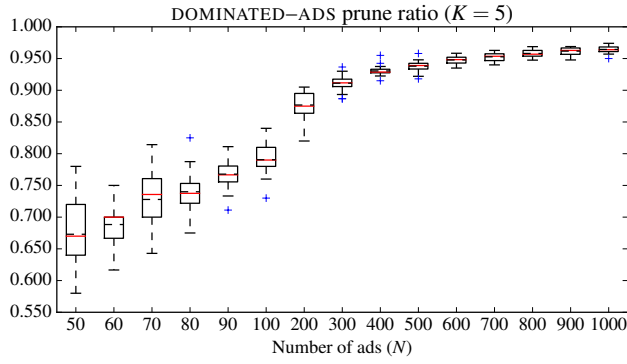
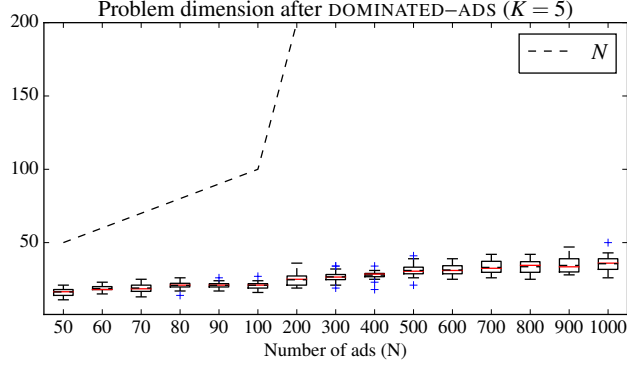
Time–approximation tradeoffs



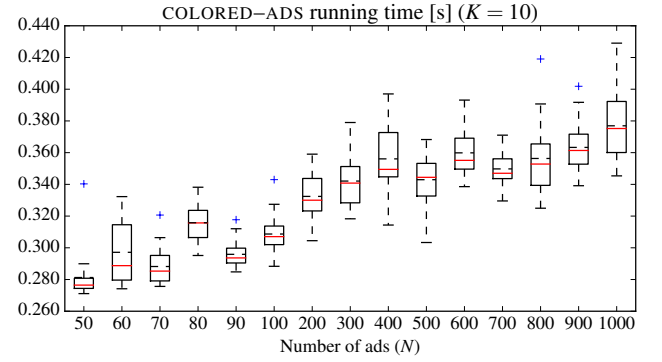
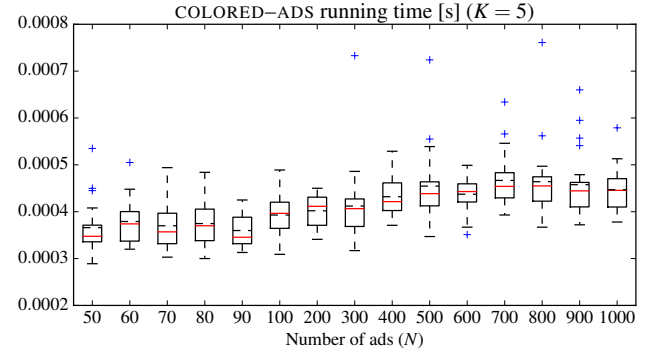
Experimental evaluation

We prove the statistical significance of the experimental results in the following boxplots. We follow the structure of the paper.

DOMINATED-ADS algorithm



COLORED-ADS algorithm



SORTED-ADS algorithm

